(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2002－202789
(P2002－202789A)
(43)公開日　平成14年７月19日(2002.7.19)

(51)Int.Cl.⁷	識別記号	ＦＩ		テーマコード(参考)
Ｇ１０Ｌ　13/06		Ｇ１０Ｌ	5/04	Ｆ　５Ｄ０４５
13/08			3/00	Ｈ
13/04			5/02	Ｋ

審査請求　未請求　請求項の数９　ＯＬ　（全 12 頁）

(21)出願番号　特願2000－400788(P2000－400788)

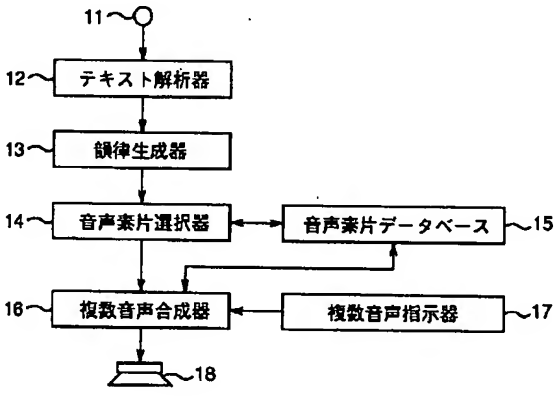
(22)出願日　平成12年12月28日(2000.12.28)

(71)出願人　000005049
シャープ株式会社
大阪府大阪市阿倍野区長池町22番22号
(72)発明者　森尾　智一
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内
(72)発明者　木村　治
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内
(74)代理人　100062144
弁理士　青山　葆　（外１名）

Ｆターム(参考)　5D045 AA07

(54)【発明の名称】　テキスト音声合成装置およびプログラム記録媒体

(57)【要約】
【課題】　簡単な処理で同一テキストを複数の話者に同時に発声させる。
【解決手段】　複数音声指示器１７は、複数音声合成器１６に対して、ピッチの変形率と混合率とを指示する。複数音声合成器１６は、音声素片データベース１５から読み出された音声素片データと音声素片選択器１４からの韻律情報とに基づいて波形重畳によって標準音声信号を生成する。さらに、上記韻律情報と複数音声指示器１７からの指示情報とに基づいて、上記標準音声信号の時間軸を伸縮して声の高さを変える。そして、上記標準音声信号と伸縮音声信号とを混合して出力端子１８から出力する。したがって、テキスト解析や韻律生成の処理を時分割で並行して行ったり、ピッチ変換処理を後処理として加えることなく、同一のテキストに基づく複数話者による同時発声を実現できる。



【特許請求の範囲】
【請求項１】　入力されたテキスト情報の読みおよび品詞情報に基づいて音声素片データベースから必要な音声素片情報を選択し、この選択された音声素片情報に基づいて音声信号を生成するテキスト音声合成装置において、
上記入力テキスト情報を解析して読みおよび品詞情報を得るテキスト解析手段と、
上記読みおよび品詞情報に基づいて韻律情報を生成する韻律生成手段と、
同一の入力テキストに対する複数音声の同時発声を指示する複数音声指示手段と、
上記複数音声指示手段からの指示を受け、上記韻律生成手段からの韻律情報と上記音声素片データベースから選択された音声素片情報とに基づいて、複数の合成音声信号を生成する複数音声合成手段を備えたことを特徴とするテキスト音声合成装置。
【請求項２】　請求項１に記載のテキスト音声合成装置において、
上記複数音声合成手段は、
上記音声素片情報と韻律情報とに基づいて、波形重畳法によって音声信号を生成する波形重畳手段と、
上記韻律情報と上記複数音声指示手段からの指示情報とに基づいて、上記波形重畳手段によって生成された音声信号の波形の時間軸を伸縮して声の高さが異なる音声信号を生成する波形伸縮手段と、
上記波形重畳手段からの音声信号と上記波形伸縮手段からの音声信号とを混合する混合手段を備えていることを特徴とするテキスト音声合成装置。
【請求項３】　請求項１に記載のテキスト音声合成装置において、
上記複数音声合成手段は、
上記音声素片情報と韻律情報とに基づいて、波形重畳法によって音声信号を生成する第１波形重畳手段と、
上記音声素片情報と韻律情報と上記複数音声指示手段からの指示情報とに基づいて、上記第１波形重畳手段とは異なる基本周期で、上記波形重畳法によって音声信号を生成する第２波形重畳手段と、
上記第１波形重畳手段からの音声信号と上記第２波形重畳手段からの音声信号とを混合する混合手段を備えていることを特徴とするテキスト音声合成装置。
【請求項４】　請求項１に記載のテキスト音声合成装置において、
上記複数音声合成手段は、
上記音声素片情報と韻律情報とに基づいて、波形重畳法によって音声信号を生成する第１波形重畳手段と、
上記音声素片データベースとしての第１音声素片データベースとは異なる音声素片情報が格納された第２音声素片データベースと、
上記２音声素片データベースから選択された音声素片情報と、上記韻律情報と、上記複数音声指示手段からの指示情報とに基づいて、上記波形重畳法によって音声信号を生成する第２波形重畳手段と、
上記第１波形重畳手段からの音声信号と上記第２波形重畳手段からの音声信号とを混合する混合手段を備えていることを特徴とするテキスト音声合成装置。
【請求項５】　請求項１に記載のテキスト音声合成装置において、
上記複数音声合成手段は、
上記音声素片と韻律情報とに基づいて、波形重畳法によって音声信号を生成する波形重畳手段と、
上記韻律情報と上記複数音声指示手段からの指示情報とに基づいて上記音声素片の波形の時間軸を伸縮し、上記波形重畳法によって音声信号を生成する波形伸縮重畳手段と、
上記波形重畳手段からの音声信号と上記波形伸縮重畳手段からの音声信号とを混合する混合手段を備えていることを特徴とするテキスト音声合成装置。
【請求項６】　請求項１に記載のテキスト音声合成装置において、
上記複数音声合成手段は、
上記韻律情報に基づいて、第１励振波形を生成する第１励振波形生成手段と、
上記韻律情報と上記複数音声指示手段からの指示情報とに基づいて、上記第１励振波形とは周波数が異なる第２励振波形を生成する第２励振波形生成手段と、
上記第１励振波形と第２励振波形とを混合する混合手段と、
上記音声素片情報に含まれている声道調音特性パラメータを取得し、この声道調音特性パラメータを用いて、上記混合された励振波形に基づいて合成音声信号を生成する合成フィルタを備えていることを特徴とするテキスト音声合成装置。
【請求項７】　請求項２乃至請求項６の何れか一つに記載のテキスト音声合成装置において、
上記波形伸縮手段，第２波形重畳手段，波形伸縮重畳手段あるいは第２励振波形生成手段は、複数存在することを特徴とするテキスト音声合成装置。
【請求項８】　請求項２乃至請求項７の何れか一つに記載のテキスト音声合成装置において、
上記混合手段は、上記複数音声指示手段からの指示情報に基づく混合率で上記混合を行うようになっていることを特徴とするテキスト音声合成装置。
【請求項９】　コンピュータを、
請求項１におけるテキスト解析手段，韻律生成手段，複数音声指示手段および複数音声合成手段として機能させるテキスト音声合成処理プログラムが記録されたことを特徴とするコンピュータ読出し可能なプログラム記録媒体。
【発明の詳細な説明】

【0001】
【発明の属する技術分野】この発明は、テキストから合成音声信号を生成するテキスト音声合成装置およびテキスト音声合成処理プログラムを記録したプログラム記録媒体に関する。
【0002】
【従来の技術】図11は、一般的なテキスト音声合成装置の構成を示すブロック図である。テキスト音声合成装置は、テキスト入力端子1,テキスト解析器2,韻律生成器3,音声素片選択器4,音声素片データベース5,音声合成器6および出力端子7で概略構成される。
【0003】以下、従来のテキスト音声合成装置の動作について説明する。入力端子1から単語や文章等の日本語の漢字仮名混じりテキスト情報(例えば、漢字「左」)が入力されると、テキスト解析器2は、入力テキスト情報「左」を読みの情報（例えば、「hidari」）に変換して出力する。尚、入力テキストとしては、日本語の漢字仮名混じりテキストに限定されるものではなく、アルファベット等の読み記号を直接入力しても差し支えない。
【0004】上記韻律生成器3は、上記テキスト解析器2からの読み情報「hidari」に基づいて、韻律情報(声の高さ,大きさ,発声速度の情報)を生成する。ここで、声の高さの情報は母音のピッチ(基本周波数)で設定され、本例の場合においては、時間順に母音「i」,「a」,「i」のピッチが設定される。また、声の大きさおよび発声速度の情報は、各音素「h」,「i」,「d」,「a」,「r」,「i」毎に音声波形の振幅および継続時間長で設定される。こうして生成された韻律情報は、読み情報「hidari」と共に音声素片選択器4に送出される。
【0005】そうすると、上記音声素片選択器4は、音声素片データベース5を参照して、韻律生成器3からの読み情報「hidari」に基づいて音声合成に必要な音声素片データを選択する。ここで、音声合成単位としては、子音+母音(CV：Consonant, Vowel)の音節単位(例えば「ka」,「gu」)や、高音質化を目的に音素連鎖の過渡部の特徴量を保持した母音+子音+母音(VCV)の単位(例えば「aki」,「ito」)等が広く用いられている。以下の説明においては、音声素片の基本単位(音声合成単位)としてVCV単位を用いる場合について説明する。
【0006】上記音声素片データベース5には、例えばアナウンサーの発声した音声データからVCVの単位で適切に切り出された音声データを分析し、合成処理に必要な形式に変換された波形やパラメータが、上記音声素片データとして格納されている。VCV音声素片を合成単位として用いる一般的な日本語テキスト音声合成の場合には、800個程度のVCV音声素片データが格納されている。本例のごとく読み情報「hidari」が音声素片選択器4に入力された場合には、音声素片選択器4は、音声素片データベース5から、VCV素片「*hi」,「idi」,「ari」,「i**」の音声素片データを選択するのである。尚、記号「*」は無音を表す。こうして得られた選択結果情報は、韻律情報と共に音声合成器6に送出される。
【0007】最後に、上記音声合成器6は、入力された選択結果情報に基づいて音声素片データベース5から該当する音声素片データを読み出す。そして、入力された韻律情報と上記得られた音声素片データとに基づいて、韻律情報に従って声の高さや大きさや発声速度を制御しながら、上記選択されたVCV音声素片の系列を母音区間で滑らかに接続して、出力端子7から出力するのである。ここで、上記音声合成器6には、一般に波形重畳方式と呼ばれる手法(例えば、特開昭60-21098号公報)や、一般にボコーダー方式またはホルマント合成方式と呼ばれる手法(例えば、「音声情報処理の基礎」オーム社P76-77)が広く用いられている。
【0008】上記テキスト音声合成装置は、声の高さや音声素片データベースを変更することによって、声質(話者)を増やすことができる。また、上記音声合成器6からの出力音声信号に対して別途信号処理を行うことによって、エコー等の音響効果を施すことも行われている。さらに、音声合成器6からの出力音声信号に対してカラオケ等にも応用されているピッチ変換処理を施し、元々の合成音声信号とピッチ変換音声信号とを組み合わせて複数話者の同時発声を行うことが提案されている(例えば、特開平3-211597号公報)。また、上記テキスト音声合成装置におけるテキスト解析器2および韻律生成器3を時分割で駆動すると共に、音声合成器6等によって構成される音声出力部を複数設けることによって、複数のテキストに対する複数の音声を同時に出力する装置も提案されている(例えば、特開平6-75594号公報)。
【0009】
【発明が解決しようとする課題】しかしながら、上記従来のテキスト音声合成装置においては、音声素片データベースを変更することによって、指定したテキストを種々の話者に切り替えて発声することは可能ではある。ところが、例えば、同一内容を複数人で同時に発声させることは不可能であるという問題がある。
【0010】また、上記特開平6-75594号公報に開示されているように、上記テキスト音声合成装置におけるテキスト解析器2および韻律生成器3を時分割で駆動すると共に、上記音声出力部を複数設けることによって、複数の合成音声を同時に出力することができる。しかしながら、時分割で前処理を行う必要があり、装置が複雑化すると言う問題がある。
【0011】また、上記特開平3-211597号公報に開示されているように、上記音声合成器6からの出力音声信号に対してピッチ変換処理を施して、標準の合成音声信号とピッチ変換音声信号とによって複数話者を同

時発声させることができる。しかしながら、上記ピッチ変換処理には、一般にピッチ抽出と言われる処理量の大きい処理が必要であり、そのような装置構成では処理量が多くなると共にコストの増加も大きいと言う問題がある。

【0012】そこで、この発明の目的は、より簡単な処理で同一テキストを複数の話者に同時に発声させることが可能なテキスト音声合成装置、および、テキスト音声合成処理プログラムを記録したプログラム記録媒体を提供することにある。

【0013】

【課題を解決するための手段】上記目的を達成するため、第1の発明は、入力されたテキスト情報の読み及び品詞情報に基づいて音声素片データベースから必要な音声素片情報を選択し,この選択された音声素片情報に基づいて音声信号を生成するテキスト音声合成装置において、上記入力テキスト情報を解析して読みおよび品詞情報を得るテキスト解析手段と、上記読みおよび品詞情報に基づいて韻律情報を生成する韻律生成手段と、同一の入力テキストに対する複数音声の同時発声を指示する複数音声指示手段と、上記複数音声指示手段からの指示を受け,上記韻律生成手段からの韻律情報と上記音声素片データベースから選択された音声素片情報とに基づいて,複数の合成音声信号を生成する複数音声合成手段を備えたことを特徴としている。

【0014】上記構成によれば、一つのテキスト情報からテキスト解析手段および韻律生成手段によって読みおよび韻律情報が生成される。そして、複数音声指示手段からの指示に従って、複数音声合成手段によって、上記一つのテキスト情報から生成された韻律情報と音声素片データベースから選択された音声素片情報とに基づいて複数の合成音声信号が生成される。したがって、同一の入力テキストに基づく複数音声の同時発声が、テキスト解析手段および韻律生成手段の時分割処理やピッチ変換処理の追加等を行うことなく簡単な処理で行われる。

【0015】また、第1の実施例は、上記複数音声合成手段を、上記音声素片情報と韻律情報とに基づいて,波形重畳法によって音声信号を生成する波形重畳手段と、上記韻律情報と上記複数音声指示手段からの指示情報とに基づいて,上記波形重畳手段によって生成された音声信号の波形の時間軸を伸縮して声の高さが異なる音声信号を生成する波形伸縮手段と、上記波形重畳手段からの音声信号と上記波形伸縮手段からの音声信号とを混合する混合手段を備えるように成したことを特徴としている。

【0016】この実施例によれば、波形重畳手段によって、標準の音声信号が生成される。一方、波形伸縮手段によって、上記標準の音声信号の波形の時間軸が伸縮されて伸縮音声信号が生成される。そして、混合手段によって、上記標準の音声信号と伸縮音声信号とが混合される。こうして、例えば、同一の入力テキストに基づく男性の音声と女性の音声とが、同時に発声される。

【0017】また、第2の実施例は、上記複数音声合成手段を、上記音声素片情報と韻律情報とに基づいて,波形重畳法によって音声信号を生成する第1波形重畳手段と、上記音声素片情報と韻律情報と上記複数音声指示手段からの指示情報とに基づいて,上記第1波形重畳手段とは異なる基本周期で,上記波形重畳法によって音声信号を生成する第2波形重畳手段と、上記第1波形重畳手段からの音声信号と上記第2波形重畳手段からの音声信号とを混合する混合手段を備えるように成したことを特徴としている。

【0018】この実施例によれば、第1波形重畳手段によって、上記音声素片に基づいて第1の音声信号が生成される。一方、第2波形重畳手段によって、上記音声素片に基づいて上記第1の音声信号とは基本周期のみが異なる第2の音声信号が生成される。そして、混合手段によって、上記第1の音声信号と第2の音声信号とが混合される。こうして、例えば、同一の入力テキストに基づく男性の音声と男性の更に高音の音声とが、同時に発声される。

【0019】さらに、上記第1波形重畳手段と第2波形重畳手段との基本構成は同じであるため、1つの波形重畳手段を時分割によって上記第1波形重畳手段と第2波形重畳手段として動作させることが可能であり、構成を簡単にして低コスト化を図ることが可能になる。

【0020】また、第3の実施例は、上記複数音声合成手段を、上記音声素片情報と韻律情報とに基づいて,波形重畳法によって音声信号を生成する第1波形重畳手段と、上記音声素片データベースとしての第1音声素片データベースとは異なる音声素片情報が格納された第2音声素片データベースと、上記第2音声素片データベースから選択された音声素片情報と,上記韻律情報と,上記複数音声指示手段からの指示情報とに基づいて,上記波形重畳法によって音声信号を生成する第2波形重畳手段と、上記第1波形重畳手段からの音声信号と上記第2波形重畳手段からの音声信号とを混合する混合手段を備えるように成したことを特徴としている。

【0021】この実施例によれば、例えば、第1音声素片データベースに男性用の音声素片情報を格納する一方、第2音声素片データベースに女性用の音声素片情報を格納しておけば、上記第2波形重畳手段は上記第2音声素片データベースから選択された音声素片情報を用いることによって、同一の入力テキストに基づく男性の音声と女性の音声とが、同時に発声される。

【0022】また、第4の実施例は、上記複数音声合成手段を、上記音声素片と韻律情報とに基づいて,波形重畳法によって音声信号を生成する波形重畳手段と、上記韻律情報と上記複数音声指示手段からの指示情報とに基づいて上記音声素片の波形の時間軸を伸縮し,上記波形

重畳法によって音声信号を生成する波形伸縮重畳手段と、上記波形重畳手段からの音声信号と上記波形伸縮重畳手段からの音声信号とを混合する混合手段を備えるように成したことを特徴としている。

【００２３】この実施例によれば、波形重畳手段によって、上記音声素片が用いられて標準の音声信号が生成される。一方、波形伸縮重畳手段によって、上記音声素片の波形の時間軸が伸縮されて、上記標準の音声信号とはピッチが異なり且つ周波数スペクトルが変形された音声信号が生成される。そして、混合手段によって、上記両音声信号が混合される。こうして、例えば、同一の入力テキストに基づく男性の音声と女性の音声とが、同時に発声される。

【００２４】また、第５の実施例は、上記複数音声合成手段を、上記韻律情報に基づいて,第１励振波形を生成する第１励振波形生成手段と、上記韻律情報と上記複数音声指示手段からの指示情報とに基づいて,上記第１励振波形とは周波数が異なる第２励振波形を生成する第２励振波形生成手段と、上記第１励振波形と第２励振波形とを混合する混合手段と、上記音声素片情報に含まれている声道調音特性パラメータを取得し,この声道調音特性パラメータを用いて,上記混合された励振波形に基づいて合成音声信号を生成する合成フィルタを備えるように成したことを特徴としている。

【００２５】この実施例によれば、第１励振波形生成手段によって生成された第１励振波形と第２励振波形生成手段によって生成された上記第１励振波形とは周波数が異なる第２励振波形との混合励振波形が、混合手段によって生成される。そして、この混合励振波形に基づいて、上記選択された音声素片情報に含まれる声道調音特性パラメータによって声道調音特性が設定された合成フィルタによって、合成音声が生成される。こうして、例えば、同一の入力テキストに基づく複数の声の高さの音声が、同時に発声される。

【００２６】また、第６の実施例は、上記波形伸縮手段,第２波形重畳手段,波形伸縮重畳手段あるいは第２励振波形生成手段を、複数設けたことを特徴としている。

【００２７】この実施例によれば、同一の入力テキストに基づいて同時発声させる際の人数を３人以上に増加でき、バラエティーに富んだテキスト合成音声が生成される。

【００２８】また、第７の実施例は、上記混合手段を、上記複数音声指示手段からの指示情報に基づく混合率で上記混合を行うように成したことを特徴としている。

【００２９】この実施例によれば、同一の入力テキストに基づいて同時発声させる複数の人夫々に遠近感を持たせたりして、種々の場面に応じた複数人による同時発声が可能になる。

【００３０】また、第２の発明のプログラム記録媒体は、コンピュータを、上記第１の発明におけるテキスト解析手段,韻律生成手段,複数音声指示手段および複数音声合成手段として機能させるテキスト音声合成処理プログラムが記録されたことを特徴としている。

【００３１】上記構成によれば、上記第１の発明の場合と同様に、同一の入力テキストに基づく複数音声の同時発声が、テキスト解析手段および韻律生成手段の分割処理やピッチ変換処理の追加等を行うことなく簡単な処理で行われる。

【００３２】

【発明の実施の形態】以下、この発明を図示の実施の形態により詳細に説明する。

＜第１実施の形態＞図１は、本実施の形態のテキスト音声合成装置におけるブロック図である。本テキスト音声合成装置は、テキスト入力端子１１,テキスト解析器１２,韻律生成器１３,音声素片選択器１４,音声素片データベース１５,複数音声合成器１６,複数音声指示器１７および出力端子１８で概略構成される。

【００３３】上記テキスト入力端子１１,テキスト解析器１２,韻律生成器１３,音声素片選択器１４,音声素片データベース１５および出力端子１８は、図１１に示す従来のテキスト音声合成装置におけるテキスト入力端子１,テキスト解析器２,韻律生成器３,音声素片選択器４,音声素片データベース５および出力端子７と同様である。すなわち、入力端子１１から入力されたテキスト情報は、テキスト解析器１２によって読みの情報に変換される。そして、韻律生成器１３によって上記読み情報に基づいて韻律情報が生成され、音声素片選択器１４によって、音声素片データベース１５から上記読み情報に基づいてＶＣＶ音声素片が選択され、選択結果情報が韻律情報と共に複数音声合成器１６に送出されるのである。

【００３４】上記複数音声指示器１７は、上記複数音声合成器１６に対してどのような複数の音声を同時に発声するのかを指示する。そうすると、複数音声合成器１６は、複数音声指示器１７からの指示に従って複数の音声信号を同時に合成するのである。そうすることによって、同一の入力テキストに基づいて複数の話者によって同時に発声させることができるのである。例えば、「いらっしゃいませ」という発声を、男声と女声との２名の話者で同時に行うことが可能になるのである。

【００３５】上記複数音声指示器１７は、上述したように、上記複数音声合成器１６に対して、どのような複数の声で発声させるかを指示する。その場合の指示の例としては、通常の合成音声に対するピッチの変化率と、ピッチを変化させた音声信号の混合率とを指定する方法がある。例えば「１オクターブ上の音声信号を、振幅を半分にして混合する」という指定である。尚、上述の例では、２つの音声を同時に発声させる例で説明しているが、処理量やデータベースのサイズの増加は生じるものの、３つ以上の音声の同時発声にも容易に拡張できる。

【００３６】上記複数音声合成器１６は、上記複数音声

指示器１７からの指示に従って、複数の音声を同時に発声させる処理を行う。後に説明するように、この複数音声合成器１６は図１１に示す１つの音声を発声させる従来のテキスト音声合成装置における音声合成器６の処理を部分的に拡充して実現することができる。したがって、上記特開平３－２１１５９７号公報の場合のようにピッチ変換処理を後処理として加える構成に比べて、複数音声生成の処理量の増加を少なく抑えることができるのである。

【００３７】以下、上記複数音声合成器１６の構成および動作について具体的に説明する。図２は、複数音声合成器１６の構成の一例を示すブロック図である。図２において、複数音声合成器１６は、波形重畳器２１,波形伸縮器２２および混合器２３から構成される。上記波形重畳器２１は、音声素片選択器１４によって選択された音声素片データを読み出し、この音声素片データと音声素片選択器１４からの韻律情報とに基づいて、波形重畳によって音声信号を生成する。そして、生成された音声信号は、波形伸縮器２２と混合器２３とに送出される。そうすると、波形伸縮器２２は、音声素片選択器１４からの韻律情報と複数音声指示器１７からの上記指示とに基づいて、波形重畳器２１からの音声信号の波形の時間軸を伸縮して声の高さを変える。そして、伸縮後の音声信号が混合器２３に送出される。混合器２３は、波形重畳器２１からの標準の音声信号と波形伸縮器２２からの伸縮後の音声信号との二つの音声信号を混合して、出力端子１８に出力するのである。

【００３８】上記構成において、上記波形重畳器２１で合成音を生成する処理としては、例えば、特開昭６０－２１０９８号公報に開示されている波形重畳方式を用いている。この波形重畳方式においては、音声素片データベース１５内に音声素片を基本周期単位の波形として記憶している。そして、波形重畳器２１は、この波形を指定のピッチに応じた時間間隔で繰り返し生成することによって音声信号を生成するのである。波形重畳の処理として種々の実現方法が開発されているが、例えば繰り返す時間間隔が音声素片の基本周波数より長い場合は不足している部分に０のデータを埋め、逆に短い場合は波形の終端が急峻に変化しないように適当に窓掛け処理を行った後に処理を打ち切る方法等がある。

【００３９】次に、上記波形伸縮器２２によって行われる上記波形重畳方式で生成された標準の音声信号による声の高さを変える処理について説明する。ここで、声の高さを変える処理は、上記特開平３－２１１５９７号公報等に開示された従来の技術においてはテキスト音声合成の出力信号に対して行うため、ピッチ抽出処理が必要である。これに対して、本実施の形態においては、複数音声合成器１６に入力される韻律情報に含まれるピッチ情報を用いるために、ピッチ抽出処理を省くことができ効率的に実現できるのである。

【００４０】図３は、本実施の形態における上記複数音声合成器１６の各部で生成される音声信号波形を示す。以下、図３に従って、声の高さを変える処理について説明する。図３(a)は、波形重畳器２１によって上記波形重畳方式で生成された母音区間の音声波形である。波形伸縮器２２は、音声素片選択器１４からの韻律情報の１つであるピッチと、複数音声指示器１７から指示されたピッチ変化率の情報とに基づいて、波形重畳器２１で生成された図３(a)の音声波形を基本周期Ａ毎に波形伸縮する。その結果、図３(b)に示すように、全体が時間軸方向に伸縮された音声波形が得られる。その際に、上記伸縮によって全体の時間長が変化しないように、ピッチを高くする場合には適当に基本周期単位の波形を多く繰り返し、逆にピッチを低くする場合には間引くようにする。図３(b)の場合には基本周期を狭めた波形に縮めているので、図３(a)の音声波形に比べピッチが高くなり、周波数スペクトルも高域に伸張された信号となる。効果を分かり易く例で説明すると、上記標準の音声信号としての男声の合成音声信号に基づいて、波形伸縮器２２によって上記伸縮された音声信号としての女声の合成音声信号が作成されたことになるのである。

【００４１】次に、上記混合器２３は、上記複数音声指示器１７から与えられる混合率に従って、波形重畳器２１で生成された図３(a)の音声波形と波形伸縮器２２で生成された図３(b)の音声波形との２つの音声波形を混合する。図３(c)に混合された結果の音声波形の一例を示す。こうして、同一のテキストに基づいて二人の話者による同時発声が実現されるのである。

【００４２】上述したごとく、本実施の形態においては、上記複数音声合成器１６と複数音声指示器１７とを有している。さらに、複数音声合成器１６を波形重畳器２１,波形伸縮器２２および混合器２３で構成している。そして、複数音声指示器１７によって、複数音声合成器１６に対して、標準の合成音声信号に対するピッチの変化率(ピッチ変化率)と、ピッチを変化させた音声信号の混合率とを指示する。

【００４３】そうすると、上記波形重畳器２１は、音声素片データベース１５から読み出された音声素片データと音声素片選択器１４からの韻律情報に基づいて、波形重畳によって標準音声信号を生成する。一方、波形伸縮器２２は、音声素片選択器１４からの韻律情報と複数音声指示器１７からの上記指示とに基づいて、上記標準の音声信号の波形の時間軸を伸縮して声の高さを変える。そして、混合器２３によって、波形重畳器２１からの標準の音声信号と波形伸縮器２２からの伸縮音声信号とを混合して、出力端子１８に出力するようにしている。

【００４４】したがって、上記テキスト解析器１２および韻律生成器１３は、時分割処理を行うことなく１つの入力テキスト情報に対してテキスト解析処理と韻律生成処理とを行えばよい。また、複数音声合成器１６の後処

理として、ピッチ変換処理を加える必要もない。すなわち、本実施の形態によれば、同一のテキストに基づく複数話者による合成音声の同時発声を、より簡単な処理で、より簡単な装置で実現することができるのである。

【００４５】<第２実施の形態>以下、上記複数音声合成器１６の他の実施の形態について説明する。図４は、本実施の形態における複数音声合成器１６の構成を示すブロック図である。本複数音声合成器１６は、第１波形重畳器２５,第２波形重畳器２６および混合器２７で構成されている。第１波形重畳器２５は、音声素片データベース１５から読み出された音声素片データと音声素片選択器１４からの韻律情報とに基づいて、上記波形重畳によって音声信号を生成して混合器２７に送出する。一方、第２波形重畳器２６は、音声素片選択器１４からの韻律情報の１つであるピッチを複数音声指示器１７から指示されたピッチ変化率に基づいて変更する。そして、第１波形重畳器２５が用いた音声素片データと同一の音声素片データと上記変更後のピッチとに基づいて、上記波形重畳によって音声信号を生成する。そして、生成した音声信号を混合器２７に送出するのである。混合器２７は、第１波形重畳器２５からの標準の音声信号と第２波形重畳器２６からの音声信号との二つの音声信号を、複数音声指示器１７からの混合率に従って混合して出力端子１８に出力するのである。

【００４６】尚、上記第１波形重畳器２５による合成音声生成処理は、上記第１実施の形態における波形重畳器２１の場合と同じである。また、上記第２波形重畳器２６による合成音声生成処理も、複数音声指示器１７からのピッチ変化率の指示に従ってピッチを変更する点を除けば、波形重畳器２１の場合と同じ通常の波形重畳処理である。したがって、上記第１実施の形態における複数音声合成器１６の場合には、波形重畳器２１とは構成を異にする波形伸縮器２２を有しているため、指定の基本周期に波形を伸縮する処理が別途必要であるのに対して、本実施の形態においては、基本の機能が同じ二つの波形重畳器２５,２６を用いるので、実際の構成においては、第１波形重畳器２５を時分割処理で２回使用することによって第２波形重畳器２６を削除することも可能であり、構成を簡単にしてコストを低減することも可能なのである。

【００４７】図５は、本実施の形態における各部で生成される音声信号波形を示す。以下、図５に従って音声信号生成処理について説明する。図５(a)は、第１波形重畳器２５によって標準の波形重畳方式で生成された母音区間の音声波形である。図５(b)は、第２波形重畳器２６によって、複数音声指示器１７から指示されたピッチ変化率に基づいて変更したピッチを用いて、標準のピッチとは異なるピッチで生成された音声波形である。この例では通常より高いピッチの音声信号が生成されている。尚、図５(b)から分かるように、第２波形重畳器２６によって生成された音声信号は、図５(a)の音声波形に対してピッチは変化しているが波形伸縮は行われないので、周波数スペクトルは第１波形重畳器２５による標準の音声波形と同じである。効果を分かり易く例で説明すると、上記標準の音声信号としての男声の合成音声信号に基づいて、第２重畳器２６によってピッチを高めた男声の合成音声信号が作成されたことになるのである。

【００４８】次に、上記混合器２７は、上記複数音声指示器１７から与えられる混合率に従って、第１波形重畳器２５で生成された図５(a)の音声波形と第２波形重畳器２６で生成された図５(b)の音声波形との２つの音声波形を混合する。図５(c)に混合された結果の音声波形の一例を示す。こうして、同一のテキストに基づいて二人の話者による同時発声が実現されるのである。

【００４９】上述したごとく、本実施の形態においては、上記複数音声合成器１６を第１波形重畳器２５,第２波形重畳器２６および混合器２７で構成している。そして、第１波形重畳器２５によって、音声素片データベース１５から読み出された音声素片データに基づいて標準の音声信号を生成する。一方、第２波形重畳器２６によって、音声素片選択器１４からのピッチを複数音声指示器１７からのピッチ変化率に基づいて変更したピッチを用いて、上記音声素片データに基づいて上記波形重畳によって音声信号を生成する。そして、混合器２７によって、両波形重畳器２５,２６からの二つの音声信号を混合して、出力端子１８に出力するようにしている。したがって、同一のテキストに基づいて二人の話者による同時発声を簡単な処理で行うことができるのである。

【００５０】また、本実施の形態によれば、基本の機能が同じ二つの波形重畳器２５,２６を用いるので、第１波形重畳器２５を時分割処理で２回使用することによって第２波形重畳器２６を削除することも可能であり、上記第１実施の形態に比して、構成を簡単にしてコスト低減を図ることが可能になる。

【００５１】<第３実施の形態>図６は、本実施の形態における複数音声合成器１６の構成を示すブロック図である。本複数音声合成器１６は、波形重畳器３１,波形伸縮重畳器３２及び混合器３３で構成されている。波形重畳器３１は、音声素片データベース１５から読み出された音声素片データと音声素片選択器１４からの韻律情報とに基づいて、上記波形重畳によって音声信号を生成して混合器３３に送出する。一方、波形伸縮重畳器３２は、音声素片データベース１５から読み出された波形重畳器３１が用いた音声素片データと同じ音声素片の波形を、複数音声指示器１７から指示されたピッチ変化率に基づいて指定のピッチに応じた時間間隔に伸縮して繰り返し生成することによって音声信号を生成する。その場合における上記伸縮の方法としては、線形補間等がある。すなわち、本実施の形態においては、波形重畳器自体に波形伸縮機能を持たせて波形重畳の処理過程におい

て音声素片の波形を伸縮するのである。

【００５２】こうして生成された音声信号は混合器３３に送出される。そうすると、混合器２８は、波形重畳器３１からの標準の音声信号と波形伸縮重畳器３２からの伸縮音声信号との二つの音声信号を、複数音声指示器１７から与えられた混合率に従って混合し、出力端子１８に出力するのである。

【００５３】本実施の形態の複数音声合成器１６における上記波形重畳器３１,波形伸縮重畳器３２および混合器３３よって生成される音声信号の波形は、図３と同様である。尚、上記第２実施の形態における第２波形重畳器２６から出力される音声信号もピッチは変化しているが、周波数スペクトルは変化していないので、声質的には似ている複数の声が出力される。これに対して、本実施の形態における波形伸縮重畳器３２から出力される音声信号は、周波数スペクトルも変化されているのである。

【００５４】＜第４実施の形態＞図７は、本実施の形態における複数音声合成器１６の構成を示すブロック図である。本複数音声合成器１６は、第２実施の形態の場合と同様に、第１波形重畳器３５,第２波形重畳器３６および混合器３７で構成されている。さらに、本実施の形態においては、第２波形重畳器３６が専用に用いる音声素片データベースを、第１波形重畳器３５が用いる音声素片データベース１５と独立して設けている。以下、第１波形重畳器３５が用いる音声素片データベース１５を第１音声素片データと称する一方、第２波形重畳器３６が用いる音声素片データベースを第２音声素片データベース３８と称する。

【００５５】上記第１実施の形態～第３実施の形態においては、ある一人の話者の声から作成された音声素片データベース１５のみを用いているが。本実施の形態においては、音声素片データベース１５とは別の話者から作成された第２音声素片データベース３８を備えて、第２波形重畳器３６によって用いられるのである。この発明の場合には、元々異なる声質の２種類の音声データベース１５,３８を用いるので、上記各実施の形態以上にバリエーションに富んだ複数の音質の同時発声が可能になる。

【００５６】尚、この場合には、上記複数音声指示器１７からは、複数の音声素片データベースを用いて複数の音声合成を行う指定が出力される。例えば「通常の合成音声の生成には男性話者のデータを用い、もう一つの合成音声の生成には別途女性話者のデータベースを用いて、二つを同比率で混合する」という指定である。

【００５７】図８は、本実施の形態における上記複数音声合成器１６の各部によって生成される音声信号波形を示す。以下、図８に従って音声信号生成処理について説明する。図８(a)は、第１音声素片データベース１５を用いて第１波形重畳器３５によって生成された標準音声波形である。また、図８(b)は、第２音声素片データベース３８を用いて第２波形重畳器３６によって生成された標準音声波形よりもピッチが高い音声信号波形である。また、図８(c)は、上記２つの音声波形を混合した音声波形である。尚、この場合、第１音声素片データベース１５を男性話者から作成する一方、第２音声素片データベース３８を女性話者から作成しておけば、第２波形重畳器３６において波形の伸縮処理は行わずに女性の音声を生成できるのである。

【００５８】＜第５実施の形態＞図９は、本実施の形態における複数音声合成器１６の構成を示すブロック図である。本複数音声合成器１６は、第１励振波形生成器４１,第２励振波形生成器４２,混合器４３および合成フィルタ４４で構成されている。第１励振波形生成器４１は、音声素片選択器１４からの韻律情報の１つのピッチに基づいて標準の励振波形を生成する。また、第２励振波形生成器４２は、上記ピッチを複数音声指示器１７から指示されたピッチ変化率に基づいて変更する。そして、この変更後のピッチに基づいて励振波形を生成する。また、混合器４３は、第１,第２励振波形生成器４１,４２からの２つの励振波形を、複数音声指示器１７からの混合率に従って混合して混合励振波形を生成する。また、合成フィルタ４４は、音声素片データベース１５からの音声素片データに含まれている声道調音特性を表現するパラメータを取得する。そして、この声道調音特性パラメータを用いて、上記混合励振波形に基づいて音声信号を生成する。

【００５９】すなわち、本複数音声合成器１６は、ボコーダー方式による音声合成処理を行うものであり、母音等の有声区間ではピッチに応じた時間間隔のパルス列で成る一方、摩擦性の子音等の無声区間では白色雑音で成る励振波形を生成する。そして、その励振波形を、選択された音声素片に応じた声道調音特性を与える合成フィルタを通すことによって合成音声信号を生成するのである。

【００６０】図１０は、本実施の形態における上記複数音声合成器１６の各部によって生成される音声信号波形を示す。以下、図１０に従って、本実施の形態における音声信号生成処理について説明する。図１０(a)は、第１励振波形生成器４１によって生成された標準の励振波形である。また、図１０(b)は、第２励振波形生成器４２によって生成された励振波形である。この例の場合には、複数音声指定器１７から指示されたピッチ変化率に基づいて、音声素片選択器１４からのピッチを変更した通常のピッチより高いピッチで生成されている。混合器４３は、複数音声指示器１７からの混合率に従って上記２つの励振波形を混合し、図１０(c)に示すような混合された励振波形を生成する。図１０(d)は、この混合励振波形を合成フィルタ４４に入力して得られた音声信号である。

【００６１】上記各実施の形態における音声素片データベース１５，３８には波形重畳用の音声素片の波形データが記憶されている。これに対して、本実施の形態におけるボコーダー方式用の上記音声素片データベース１５には、各音声素片毎に声道調音特性パラメータ(例えば、線形予測パラメータ)のデータが記憶されている。

【００６２】上述したごとく、本実施の形態においては、上記複数音声合成器１６を第１励振波形生成器４１，第２励振波形生成器４２，混合器４３および合成フィルタ４４で構成している。そして、第１励振波形生成器４１によって標準の励振波形を生成する。一方、第２励振波形生成器４２によって、音声素片選択器１４からのピッチを複数音声指示器１７からのピッチ変化率に基づいて変更したピッチを用いて励振波形を生成する。そして、混合器４３によって、両励振波形生成器４１，４２からの二つの励振波形を混合し、上記選択された音声素片に応じた声道調音特性に設定された合成フィルタ４４を通すことによって合成音声信号を生成するようにしている。

【００６３】したがって、本実施の形態によれば、上記テキスト解析処理および韻律生成処理を時分割で行ったり、ピッチ変換処理を後処理として加えることなく、同一のテキストに基づく複数話者による合成音声の同時発声を簡単な処理で実現することができるのである。

【００６４】尚、上記各実施の形態においては、摩擦性の子音等の無声区間に関しては上述の処理は行わず、一人の話者の合成音声信号のみを生成するようにしている。つまり、二人が同時に発声しているように信号処理するのはピッチが存在する有声区間のみなのである。また、上記第１実施の形態における波形伸縮器２２，第２実施の形態における第２波形重畳器２６，第３実施の形態における波形伸縮重畳器３２，第４実施の形態における第２波形重畳器３６および第５実施の形態における第２励振波形生成器４２を複数設けて、同一の入力テキストに基づいて同時発声させる際の人数を３人以上にすることもできる。

【００６５】ところで、上記各実施の形態における上記テキスト解析手段，韻律生成手段，複数音声指示手段及び複数音声合成手段としての機能は、プログラム記録媒体に記録されたテキスト音声合成処理プログラムによって実現される。上記プログラム記録媒体は、ＲＯＭ(リード･オンリ･メモリ)でなるプログラムメディアである。または、外部補助記憶装置に装着されて読み出されるプログラムメディアであってもよい。尚、何れの場合においても、上記プログラムメディアからテキスト音声合成処理プログラムを読み出すプログラム読み出し手段は、上記プログラムメディアに直接アクセスして読み出す構成を有していてもよいし、ＲＡＭ(ランダム･アクセス･メモリ)に設けられたプログラム記憶エリア(図示せず)にダウンロードして、上記プログラム記憶エリアにアクセスして読み出す構成を有していてもよい。尚、上記プログラムメディアからＲＡＭの上記プログラム記憶エリアにダウンロードするためのダウンロードプログラムは、予め本体装置に格納されているものとする。

【００６６】ここで、上記プログラムメディアとは、本体側と分離可能に構成され、磁気テープやカセットテープ等のテープ系、フロッピー（登録商標）ディスク，ハードディスク等の磁気ディスクやＣＤ(コンパクトディスク)－ＲＯＭ，ＭＯ(光磁気)ディスク，ＭＤ(ミニディスク)，ＤＶＤ(ディジタルビデオディスク)等の光ディスクのディスク系、ＩＣ(集積回路)カードや光カード等のカード系、マスクＲＯＭ，ＥＰＲＯＭ（紫外線消去型ＲＯＭ)，ＥＥＰＲＯＭ(電気的消去型ＲＯＭ)，フラッシュＲＯＭ等の半導体メモリ系を含めた、固定的にプログラムを担持する媒体である。

【００６７】また、上記各実施の形態におけるテキスト音声合成装置は、モデムを備えてインターネットを含む通信ネットワークと接続可能な構成を有していれば、上記プログラムメディアは、通信ネットワークからのダウンロード等によって流動的にプログラムを担持する媒体であっても差し支えない。尚、その場合における上記通信ネットワークからダウンロードするためのダウンロードプログラムは、予め本体装置に格納されているものとする。または、別の記録媒体からインストールされるものとする。

【００６８】尚、上記記録媒体に記録されるものはプログラムのみに限定されるものではなく、データも記録することが可能である。

【００６９】

【発明の効果】以上より明らかなように、第１の発明のテキスト音声合成装置は、テキスト解析手段で入力テキスト情報から得られた読みおよび品詞情報に基づいて、韻律生成手段によって韻律情報を生成し、複数音声指示手段からの指示があると、複数音声合成手段によって、上記韻律情報と音声素片データベースから選択された音声素片情報とに基づいて複数の合成音声信号を生成するので、同一の入力テキストに基づいて、複数の音声を同時に発声させることができる。その際に、特開平６－７５５９４号公報のごとく上記テキスト解析手段および韻律生成手段は時分割処理を行う必要がなく、特開平３－２１１５９７号公報のごとくピッチ変換処理の追加を行う必要がない。したがって、一つのテキストに基づく複数音声の同時発声を非常に簡単な処理で実現することができるのである。

【００７０】また、第１の実施例は、上記複数音声合成手段を、標準の音声信号を生成する波形重畳手段と、上記標準の音声信号の波形の時間軸を伸縮して音声信号を生成する波形伸縮手段と、上記標準の音声信号と伸縮された音声信号とを混合する混合手段で成したので、例えば、同一の入力テキストに基づく男性の音声と女性の音

声とを、簡単な処理で同時に発声させることができる。

【0071】また、第2の実施例は、上記複数音声合成手段を、標準の音声信号を生成する第1波形重畳手段と、上記第1波形重畳手段と同じ音声素片情報を用いて異なる基本周期の音声信号を生成する第2波形重畳手段と、上記標準の音声信号と基本周期が異なる音声信号とを混合する混合手段で成したので、例えば、男性の音声と男性の更に高音の音声とを、簡単な処理で同時に発声させることができる。

【0072】さらに、上記第1波形重畳手段と第2波形重畳手段との基本構成は同じであるため、1つの波形重畳手段を時分割によって上記第1波形重畳手段と第2波形重畳手段として動作させることが可能であり、構成を簡単にして低コスト化を図ることができる。

【0073】また、第3の実施例は、上記複数音声合成手段を、第1音声素片データベースから選択された音声素片情報を用いて標準の音声信号を生成する第1波形重畳手段と、少なくとも第2音声素片データベースから選択された音声素片情報を用いて異なるピッチの音声信号を生成する第2波形重畳手段と、上記標準の音声信号と異なるピッチの音声信号とを混合する混合手段で成したので、例えば、第1音声素片データベースに男性用の音声素片情報を格納する一方、第2音声素片データベースに女性用の音声素片情報を格納しておけば、同一の入力テキストに基づく男性の音声と女性の音声とを、簡単な処理で同時に発声させることができる。

【0074】また、第4の実施例は、上記複数音声合成手段を、標準の音声信号を生成する波形重畳手段と、上記波形重畳手段と同じ音声素片の波形の時間軸を伸縮して音声信号を生成する波形伸縮重畳手段と、上記波形重畳手段および波形伸縮重畳手段からの両音声信号を混合する混合手段で成したので、例えば、同一の入力テキストに基づく男性の音声と女性の音声とを、簡単な処理で同時に発声させることができる。

【0075】また、第5の実施例は、上記複数音声合成手段を、標準の第1励振波形を生成する第1励振波形生成手段と、上記第1励振波形と周波数が異なる第2励振波形を生成する第2励振波形生成手段と、上記両励振波形を混合する混合手段と、上記選択された音声素片情報に応じた声道調音特性パラメータを用いて上記混合された励振波形に基づいて合成音声信号を生成する合成フィルタで成したので、例えば、同一の入力テキストに基づいて、複数の声の高さの音声を簡単な処理で同時に発声させることができる。

【0076】すなわち、この実施例によれば、ボコーダー方式あるいはホルマント合成方式の音声合成装置においても、同一の入力テキストに基づく複数話者の音声を、簡単な処理で同時に発声させることができるのである。

【0077】また、第6の実施例は、上記波形伸縮手段,第2波形重畳手段,波形伸縮重畳手段あるいは第2励振波形生成手段を複数設けたので、同一の入力テキストに基づいて同時発声させる人数を3人以上に増加でき、バラエティーに富んだテキスト合成音声を生成することができる。

【0078】また、第7の実施例は、上記混合手段を、上記複数音声指示手段からの指示情報に基づく混合率で上記混合を行うように成したので、種々の場面に応じた複数人による同時発声が可能になる。

【0079】また、第2の発明のプログラム記録媒体は、コンピュータを、上記第1の発明におけるテキスト解析手段,韻律生成手段,複数音声指示手段および複数音声合成手段として機能させるテキスト音声合成処理プログラムが記録されているので、上記第1の発明の場合と同様に、同一の入力テキストに基づく複数音声の同時発声を、上記テキスト解析手段および韻律生成手段の分割処理やピッチ変換処理の追加等を行うことなく簡単な処理で行うことができる。

【図面の簡単な説明】

【図1】 この発明のテキスト音声合成装置におけるブロック図である。

【図2】 図1における複数音声合成器の構成の一例を示すブロック図である。

【図3】 図2に示す複数音声合成器の各部で生成される音声波形を示す図である。

【図4】 図2とは異なる複数音声合成器の構成を示すブロック図である。

【図5】 図4に示す複数音声合成器の各部で生成される音声波形を示す図である。

【図6】 図2および図4とは異なる複数音声合成器の構成を示すブロック図である。

【図7】 図2,図4および図6とは異なる複数音声合成器の構成を示すブロック図である。

【図8】 図7に示す複数音声合成器の各部で生成される音声波形を示す図である。

【図9】 図2,図4,図6および図7とは異なる複数音声合成器の構成を示すブロック図である。

【図10】 図9に示す複数音声合成器の各部で生成される信号波形を示す図である。

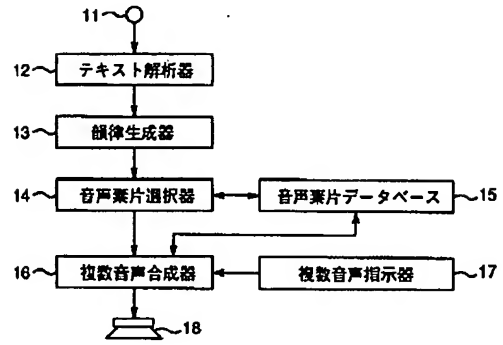
【図11】 従来のテキスト音声合成装置の構成を示すブロック図である。

【符号の説明】

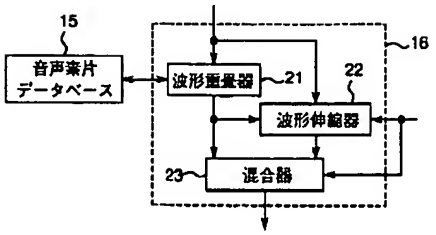
11…テキスト入力端子、
12…テキスト解析器、
13…韻律生成器、
14…音声素片選択器、
15,38…音声素片データベース、
16…複数音声合成器、
17…複数音声指示器、
18…出力端子、

21,31…波形重畳器、
22…波形伸縮器、
23,27,33,37,43…混合器、
25,35…第1波形重畳器、
26,36…第2波形重畳器、
32…波形伸縮重畳器、
41…第1励振波形生成器、
42…第2励振波形生成器、
44…合成フィルタ。

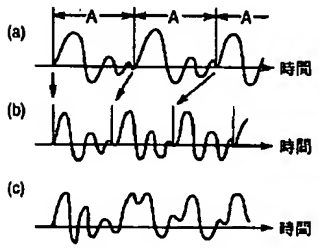
【図1】



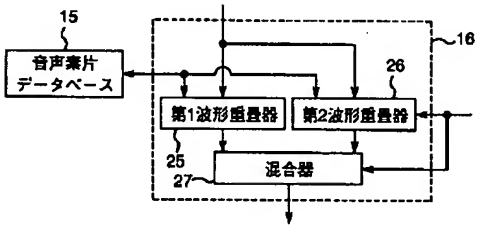
【図2】



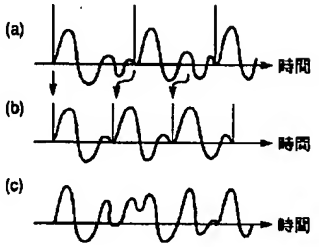
【図3】



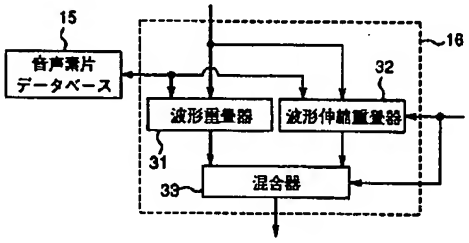
【図4】



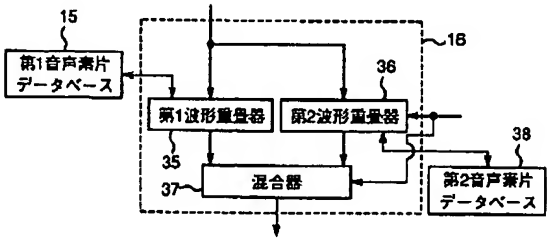
【図5】



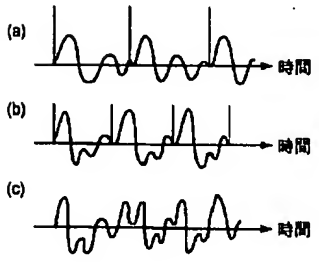
【図6】



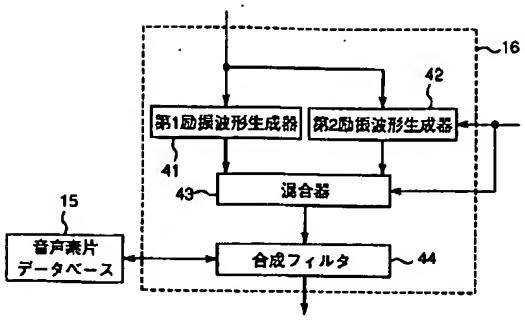
【図7】



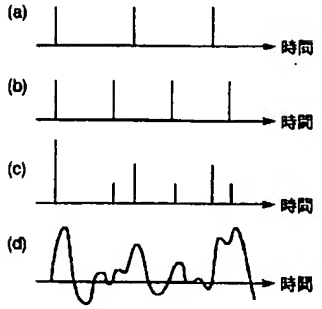
【図8】



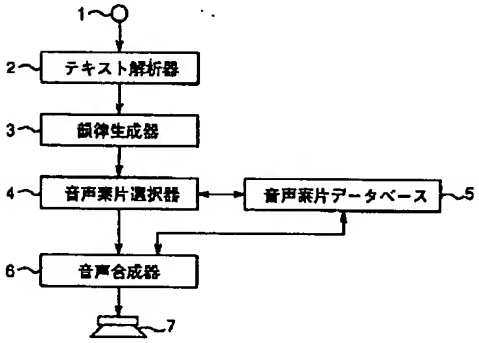
【図9】


16
42
第1励振波形生成器
第2励振波形生成器
41
43
混合器
15
音声素片
データベース
合成フィルタ
44

【図10】


(a)
時間
(b)
時間
(c)
時間
(d)
時間

【図11】


1
2
テキスト解析器
3
韻律生成器
4
音声素片選択器
音声素片データベース
5
6
音声合成器
7